

Publizieren ist nicht genug

Alexander Wagner

Deutsches Elektronen-Synchrotron DESY
Library and Documentation

Zusammenfassung

Im *Einreichungsprozess* einer Veröffentlichung erfassen Wissenschaftler traditionell die primären Metadaten ihrer Artikel, welche ggf. von Verlagen durch einige weitere Daten (Band, Seiten, DOI etc.) ergänzt werden, die eine genaue Identifikation und Zitation der jeweiligen Publikation erlauben.

Zahlreiche neue Anforderungen v. a. aus den Bereichen Reporting und Evaluation führen dazu, dass diese Verlagsmetadaten nicht mehr ausreichen um die daraus resultierenden komplexen Anfragen zu beantworten. Das führt i. d. R. zu einer erneuten Erfassung der Publikationen in speziellen Nachweissystemen (Publikationsdatenbanken) und zur Anreicherung der Metadaten. Da die Publikation auf diese Weise von deren Nachweis entkoppelt wird stellt sich bereits hier das Problem des (hinreichend) vollständigen Nachweises, resultierend in einem z. T. immensen manuellen Rechercheaufwand an den Einrichtungen, an denen diese Publikationen entstanden sind. Häufig erfolgt zusätzlich eine weitere Erfassung zur Darstellung der Inhalte in den Repositorien.

Der Nachweis in Fachdatenbanken, kommerzieller wie nichtkommerzieller Natur, erfolgt i. d. R. wiederum aus Verlagsdaten. Hier fehlen daher alle in den Publikationsdatenbanken der Einrichtungen aufwändig nachgepflegten Werte wie z. B. Projekt-, Instituts- oder Personenzuordnungen. Somit ist die Darstellung der Ergebnisse sowohl für Wissenschaftler, aber v. a. Einrichtungen und Förderer häufig nicht optimal.

Auch auf Verlagsseite führen die derzeitigen Prozesse, vor allem im Licht APC-basierter Modelle, häufig zu Problemen. So liegen Daten zur Fakturierung falsch oder unvollständig vor und eine einzeltransaktionsbasierte Abrechnung ist aufwändig und teuer.

Ausgehend von den in den unterschiedlichen Kontexten notwendigen Daten und einem optimierten Prozess wird der Frage nachgegangen, wie man durch

möglichst einfache Änderungen an den derzeitigen Prozessen einen Mehrwert für alle Beteiligten schaffen kann.

Abstract

Scientists usually generate meta data for their publications upon submission. Publishers enrich those by additional data (volume, pages, doi etc.) allowing for a precise identification and citation.

Today, new requirements especially for reporting and evaluation, require additional data associated with each publication. As the publisher generated meta data is not sufficient, this usually results in cataloguing each publication again in specialised systems (publication databases) and enrichment of the data provided by those required for the specialized needs. However, this decouples publications from reporting and easily results in a lack of completeness. Tackling this requires huge amounts of manual work at the institutions to identify own publications and add them to their respective bibliographies. Sometimes, yet another step is required to feed the publications to the institutional Open Access repositories.

Furthermore, bibliographic databases generate their content from publishers meta data. Thus, their records usually lack all the additional values from the publication databases like project assignments, affiliations or author disambiguation. From an institutional as well as a funders perspective this is far from optimal.

As soon as APCs enter the game, the current processes are a source of trouble even on the publishers end. APCs in principle require individual billing on an article basis and thus additional meta data containing funding and affiliation information. Nowadays, incomplete or even plainly wrong data result in expensive and time consuming work flows.

Starting out from a short description how articles are born today and keeping in mind the data necessary in different contexts this article investigates possible improvements. The overall goal is to suggest a properly modularized work flow that can be adopted to the current system and realize gains for all stakeholders while minimizing the necessary changes.

Einleitung: Publikationsprozess und Außendarstellung

"A mathematician is a machine for turning coffee into theorems" (Alfréd Rényi, ungarischer Mathematiker).

Hierfür benötigen, nicht nur Mathematiker, zunächst Zugriff auf die bereits publizierte Literatur ihrer Kollegen. Diese Literatur besteht heute, je nach Arbeitsfeld, nicht nur aus Artikeln und Büchern (egal welcher Medienform), sondern ebenso aus publizierten Daten oder auch der für deren Bearbeitung nötigen Software. So stellt das Sammeln relevanter Quellen einen nicht unwesentlichen Teil der wissenschaftlichen Arbeit dar, und eine gut gepflegte und strukturierte Quellensammlung ist entsprechend wertvoll. Da ferner heute Aufsätze i. d. R. mit mehreren Koautoren erstellt werden¹, muss diese Sammlung, oder zumindest die für das konkrete Projekt bzw. die geplante Publikation relevanten Teile, eben diese Kollegen zugänglich sein². Dies gilt ebenso für die gemeinsam erzielten Ergebnisse wie eigene Publikationen der Gruppe (auch Qualifikationsschriften), Daten und/oder Software, die diese Sammlung ergänzen. Da zahlreiche Arbeitsgruppen für die Verwaltung dieser Quellensammlungen lokale Nachweissysteme in Eigenregie aufbauen, mithin Ressourcen in Entwicklung und Betrieb investieren, ist die Relevanz dieser Sammlungen offensichtlich.

Zum eigentlichen Erstellen neuer Publikationen sind weitere Werkzeuge notwendig³, die ferner den Koautoren Zugriff auf die Publikation und alle ihre relevanten Teile erlauben müssen. Idealerweise steht diese mit der Quellensammlung in engem Austausch so dass ein einfaches, korrektes Zitieren möglich ist⁴.

Spätestens im Falle (oft international) verteilt arbeitender Autorentams drängt sich eine Implementation im Web auf, ist aber derzeit noch wenig

¹ Der Sonderfall des Einzelautors welcher v. a. in den Geisteswissenschaften häufig und bei Abschlussarbeiten die Regel ist, ergibt sich hier zwanglos als Grenzfall der "Ein-Personen-Kollaboration", und muss im Folgenden nicht explizit betrachtet werden.

² Es kann heute i. A. nicht davon ausgegangen werden, dass dies bei Verwendung kostenpflichtiger Angebote automatisch der Fall ist. Die Publikation im *Open Access* senkt hier offensichtliche Hürden.

³ z.B. Satzsysteme wie LaTeX oder Textverarbeitungen, Hilfsmittel zur Formatierung von Referenzen, Erstellung graphischer Darstellungen, Versionskontrollsysteme, etc.

⁴ Ein Institut am Forschungszentrum Jülich verwendet hier eine eigens entwickelte Datenbank die auf Basis von BibTeX und subversion (svn) als Versionskontrollsystem allen Koautoren die Pflege und Nutzung der Quellensammlung erlaubt.

gebräuchlich⁵. Häufig praktiziert ist der Dokumentaustausch per Email oder, institutsintern über gemeinsame Netzwerklaufwerke, und die Bearbeitung mit lokal vorhandener Software. Bei institutionsübergreifenden Kollaborationen finden (z. T. kommerzielle) cloudbasierte Speicher⁶ zunehmend Nutzer. Hier werden mittlerweile auch Web-basierte (z. T. ebenfalls kommerzielle) Lösungen zur Bearbeitung der Publikation eingesetzt⁷. All dies soll hier unter den Begriff der kollaborativen *Schreibumgebung* zusammengefasst werden. Gemein ist allen diesen Lösungen, dass sie eine (i. d. R. nicht nachnutzbare, da auf lokalen IDs basierende) *Affiliations-* und *Autorenidentifikation* implizieren, um Zugriffsrechte zu regeln⁸.

Ist nach mehreren internen Iterationen innerhalb der kollaborativen *Schreibumgebung* die von allen Beteiligten genehmigte finale Fassung (*Preprint*) entstanden wird diese heute i. d. R. über ein web-basiertes *Submission Interface* (SI) beim gewählten Publikationsorgan eingereicht. Hier werden zahlreiche Metadaten erfasst, z. T. solche, die die *Schreibumgebung* beim Erstellen des Dokuments in einer nicht-nachnutzbaren Form bereits vorliegen hatten (z. B. Autoren, Titel, Kurzfassung, Affiliationen). Ist die Publikation des Preprints, z. B. auf [arXiv](#) (Cornell University Library 2016), vorgesehen um Ergebnisse *schnell* publizieren zu können und sich so ggf. die Priorität zu sichern ist bereits hier eine Doppelerfassung von Metadaten meist unvermeidlich, da bereits hier zweimal eingereicht wird: einmal auf dem Preprint-Server und einmal beim Verlag. Um dies zumindest teilweise zu vereinfachen gehen einige Zeitschriften (z. B. JHEP) dazu über die eigentliche Submission über [arXiv](#) (Cornell University Library 2016) abzuwickeln, so dass ein Großteil der Metadaten von dort übernommen werden kann.

⁵ Angewandt wird ein solches Verfahren z. B. bei den großen LHC-Kollaborationen (ATLAS, CMS, Alice, LHC-B), die am CERN koordiniert werden. Es sei ferner erwähnt, dass derartige Prozesse bei der Erstellung von Software seit langem etabliert sind. Durch die immer stärkere Einbindung von Software in den Forschungsalltag ist eine Nutzung der dort verwendeten Werkzeuge auch zum Erstellen von Publikationen häufig naheliegend, da sich viele Wissenschaftler mit diesen für ihre tägliche Arbeit heute auseinandersetzen müssen.

⁶ z.B. [BSCW](#) (OrbiTeam Software GmbH & Co. KG 2016), [OwnCloud](#) (ownCloud 2016), [Dropbox](#) (Dropbox 2016), [Google Drive](#) (Google, Inc. 2016a)

⁷ z.B. Wikis, [Google Docs](#) (Google, Inc. 2016b), [Etherpad](#) (The Etherpad Foundation 2016) [Authorea](#) (Authorea 2016), s. a. (Perkel 2014).

⁸ Email-Adressen, Logins etc. Die immer stärkere Nutzung von [OAuth](#) (OAuth Community 2016) z. B. über [ORCID](#) (ORCID 2016) könnte perspektivisch die Wiederverwendbarkeit deutlich verbessern.

Je nach Publikationsorgan/-modell fallen, z. T. erhebliche, Kosten an⁹, deren Deckung zu klären und einzuleiten ist. Daneben entsteht auf Grund der Abrechnung pro Artikel, ein nicht zu verachtender Verwaltungsaufwand¹⁰. Erschwerend kommt die Unkenntnis vieler Wissenschaftler zu Möglichkeiten der Finanzierung der Publikationskosten hinzu. So sind die im Bereich Gold Open Access oft vorhandenen Publikationsfonds und ihre Modalitäten häufig unbekannt. Bei institutionsübergreifenden Publikationen wird ferner, insbesondere bei hohen Kosten, häufig eine Teilung angestrebt. Beides verursacht derzeit nicht unerheblichen Aufwand sowohl bei den Autoren als auch bei den beteiligten Verlagen, und wird im Kontext Open Access v. A. von letzteren gerne als Argument für die Beibehaltung des Subskriptionsmodells verwendet.

Nach Einreichung und je nach Fach und Organ verschieden ausgestalteten Zwischenschritten wie z. B. Review Prozessen, ggf. Ablehnung und Einreichen bei einer weiteren Zeitschrift, wird der Beitrag schließlich publiziert. Neben der hierdurch eintretenden Verlangsamung der Kommunikation ist der damit einhergehende Zeitverzug zwischen Ergebnis und Publikation oft gerade für junge Wissenschaftler ein Problem. Nicht selten dauert der Prozess länger als die Projektlaufzeit bzw. die für die Einreichung des Abschlussberichts vorgegebenen Fristen. So manches Projekt wird also "ohne Publikation", oder mit "in Print" bzw. "submitted" beendet. Vor allem letzterer Status, der ja dazu führen kann, dass finale Ergebnisse schlussendlich doch noch anderweitig, und mithin nicht nachvollziehbar, erscheinen ist nicht nur für Förderorganisationen unbefriedigend.

Bis vor einigen Jahren war der Publikationsprozess nach Übersendung einiger *Sonderdrucke* an die Autoren abgeschlossen.

Im Rahmen der Berichtspflicht kommen für Autoren heute jedoch zunehmend wichtiger werdende Aufgaben hinzu. Die Meldung an eine einrichtungsspezifische Publikationsdatenbank, Hochschulbibliographie oder

⁹ Bei Publikation in Closed Access Journals können dies Kosten für Überlänge (*page charges*) oder Farabbildungen (*colour charges*) sein. Im Bereich Open Access fallen ggf. *Article Processing Charges* (APC) an. Zur Diskussion diverser Kosten s. a. (Bruch u. a. 2015, Abschn. 2.2) und die dortigen Referenzen und (Björk und Solomon 2014).

¹⁰ Zur komplexen Gemengelage alleine bei Open Access-Modellen s. a. (Bruch u. a. 2015), insbesondere die Abschnitte 2.2 und 2.3 sowie die dort angegebenen Referenzen.

ein Forschungs-informationssystem (FIS) wird immer häufiger verpflichtend und dient zunehmend für *leistungsorientierte Mittelvergaben* bzw. zur Evaluation von Bereichen oder Projekten, mithin also für eine ggf. beantragte Weiterfinanzierung. Daher ist ein *vollständiger* Nachweis der eigenen Leistungen im eigenen Interesse der Wissenschaftler¹¹. Es ist unmittelbar einsichtig, dass die Berichtspflicht bereits bei nur zwei beteiligten Institutionen zur Doppelerfassung führt, und dass ferner die bereits bei der Publikation erzeugten Metadaten erneut erfasst werden müssen. Neben den reinen bibliographischen Daten (die bei Zeitschriftenaufsätzen oder Büchern häufig noch über gezielte Importe übernommen werden können) sind für die Berichte *zusätzliche* Daten erforderlich wie z. B. Projekt-, Instituts-zuordnungen, Autorenidentifikation, Rollen und Fachzuordnungen¹². Beispiele für solche Anforderungen liefert der *Kerndatensatz Forschung* (Institut für Forschungsinformation und Qualitätssicherung, Fraunhofer-Institut für Angewandte Informationstechnik und Wissenschaftsrat, Geschäftsstelle 2015) der auf Empfehlungen des Wissenschaftsrats basiert (Wissenschaftsrat 2016). In bestimmten Förderschemata sind zusätzlich die Publikationen einschließlich Volltexte dem Förderer z. T. auf vorgeschriebenem Weg zu übermitteln bzw. entsprechende Open Access-Richtlinien einzuhalten (z. B. Horizon 2020 (European Commission 2016a) und OpenAIRE (European Commission 2016b).)

Um dies leisten zu können müssen für die Meldungen zu den Nachweissystemen bibliographische Verlagsdaten z. T. erheblich mit zusätzlichen Informationen, die meist nur den Autoren bekannt sind, angereichert und derzeit häufig vielfach erfasst werden.

Gefühlt wesentlich niederschwelliger als das Reporting scheinen die Anforderungen an die eigene Außendarstellung im Internet. Im allgemeinen

¹¹ Das Eintragen von Publikationen durch zentrale Stellen, z. B. die Bibliotheken, ist zwar oft gefordert aber nicht zielführend: hier kann systembedingt nur ein Bruchteil der erbrachten Leistung erfasst werden, da nur Publikationen gefunden werden können, die bereits anderweitig nachgewiesen sind. Publikationsformen jenseits des klassischen Zeitschriftenaufsatzes und der Monographie werden in diesem Verfahren i. d. R. nur schlecht oder gar nicht berücksichtigt. Je nach Fachgebiet führt dies zu einer z. T. signifikanten systematischen Verzerrung nicht nur im Nachweis sondern ebenso in den ggf. abgeleiteten Indikatoren.

¹² In der Helmholtz-Gemeinschaft reicht z. B. auch die "einfache" Affiliationszuordnung häufig nicht aus. Publikationen müssen hier auf Teilinstitusebene abgebildet werden können. Eine Projektzuordnung ist ferner neben Drittmitteln wie z. B. EU, DFG, BMBF außerdem noch auf Basis der Helmholtz-internen sog. programmorientierten Förderung (POF) notwendig, die darüber hinaus weitere Indikatoren pro Publikation erfordert (Große 2014).

Fall kann eine Einrichtung die gewünschten Darstellungen hier nur schwer vorhersehen. Ein häufig nachgefragter Dienst sind jedoch *Publikationslisten* auf den jeweiligen Webseiten in verschiedensten Granularitäten. Einmal ist hier der persönliche Lebenslauf zu ergänzen (was automatisch nur über eine eindeutige Autorenidentifikation möglich ist), daneben die Gesamtliste des Instituts und/oder der Gruppe. Häufig sind projektbasierte Webseiten zu pflegen und bei größeren Projekten (z. B. SFB, EU-Förderung) ist die jeweilige Projektleitung über die Publikationen der Unterarbeitsgruppen zu informieren. All dies geschieht auch heute häufig manuell durch erneute Erfassung in einem bestimmten Layout, also einem unstrukturierten Anzeigeformat. Die Mehrarbeit kann hier immens sein¹³.

Anders als die periodischen Berichte wird für die Außendarstellung zusätzlich eine hohe Aktualität gefordert, die praktisch nur zu leisten ist, wenn Publikationen unmittelbar nach Einreichen auch den entsprechenden Systemen gemeldet werden. Fast alle Publikationslisten stellen daher praktisch ein *Reporting in Echtzeit* dar, d. h. beruhen auf erfassten Metadaten die nach diversen z. T. komplexen Kriterien selektiert und gruppiert werden.

Vision

Aus den vorstehenden Ausführungen ergeben sich einige Folgerungen für Infrastrukturen. Im nachfolgenden soll der Versuch unternommen werden, ausgehend von derzeit bestehenden Systemen zu signifikanten Verbesserungen zu kommen.

Sammeln, Archivieren, Strukturieren - Literaturverwaltung global

"Die Bibliothek ist eine Einrichtung, die unter archivarischen, ökonomischen und synoptischen Gesichtspunkten publizierte Information für die Benutzer sammelt, ordnet und verfügbar macht."(Ewert und Umstätter 1997)

Ausgehend von dieser Definition drängt sich auf, dass die beschriebene *Quellensammlung*, die die Grundlage wissenschaftlichen Arbeitens ist, zumindest eine große Schnittmenge mit den klassischen Aufgaben von

¹³ Dem Autor wurde hier ein "Institutsworkflow" geschildert der nicht weniger als vier unabhängige Erfassungen der gleichen Publikation und das Speichern von drei unabhängigen, verschieden benannten PDF-Kopien des gleichen Beitrags im Institut umfasst. Nicht eingerechnet die erneuten Erfassungen für notwendige Projektberichte, da die unstrukturierten Daten hier nicht nachnutzbar sind.

Bibliotheken hat. Zieht man ferner in Betracht, dass die überwiegende Mehrzahl von relevanten Publikationen in diesem Kontext textueller Natur sind und heute elektronisch zur Verfügung stehen (i. d. R. sogar *born digital*) liegt es nahe, Systeme genauer in den Blick zu nehmen, die originär geschaffen wurden um Literatur elektronisch zu archivieren und verfügbar zu machen¹⁴. Fasst man den Aufgabenbereich des, immer noch häufig als *Dissertationsserver* missbrauchten, institutionellen Repositoriums ein klein wenig weiter und ergänzt dieses System um private Bereiche, die das Sammeln von Literatur in geschlossenen Gruppen ermöglichen, kann man, leistungsfähige Schnittstellen vorausgesetzt, bereits zahlreiche Anforderungen erfüllen, die sich aus der geschilderten Sammlungsproblematik ergeben. Schließlich wurden diese Systeme geschaffen um genau dieses zu leisten: Literatur zu sammeln, zu erschließen, und dauerhaft zugänglich zu machen. Der Aspekt von privaten, von außen nicht zugänglichen Bereichen ist hier deshalb relevant, da zwar die Wissenschaft von Austausch lebt, zugleich aber kompetitiv aufgestellt ist.

In der Struktur der JOIN²-Repositorien (JOIN² 2016) ist daher der private Bereich auf Instituts- bzw. Gruppenebene eines der grundlegenden Designprinzipien (Wagner 2013, Wagner (2016)). So wird jedes eingetragene Dokument zunächst nur in diesem Bereich übernommen und erst danach ggf. weiter freigegeben. Dies gilt sowohl für die bibliographische Beschreibung als auch für ggf. vorhandene Volltexte. Grundsätzlich sollte sich jede beliebige Repository-Software so konfigurieren lassen, dass entsprechende Bereiche vorgesehen sind.

Einzig limitierender Faktor mag hier häufig die interne Verwendung des Dublin Core (Dublin Core Metadata Initiative 2016) als Datenmodell darstellen, da sich dieser nur sehr bedingt eignet um z. B. Zeitschriftenreferenzen korrekt strukturiert abzubilden und so bibliographischen Exportformaten (BibTeX, RIS etc.) sinnvoll verfügbar zu machen.

Kollaborative Texterstellung

Hat das lokale Repositorium die relevanten Funktionen wie private Bereiche und Exportschnittstellen für bibliographische Daten lassen sich dieses bereits

¹⁴ Für den Bereich der wichtiger werdenden Datensammlungen greifen analoge Überlegungen für die jeweiligen Datenrepositorien und Datenzentren.

mit der in der Einleitung angesprochenen *Schreibumgebung* verknüpfen. Die einfachste Art sind strukturierte Exporte in die jeweils verwendeten Formatierungswerkzeuge (z. B. BibTeX oder EndNote/RIS). Das ist sicherlich nicht die komfortabelste, aber die heute sofort verfügbare Möglichkeit der Einbindung des Repositoriums in eine *Schreibumgebung 1.0*. Normiert man die Schnittstelle inkl. Login-Prozedur wäre über entsprechende Web-Schnittstellen eine komplette Integration leicht möglich.¹⁵ An den am Erstellen der Publikation beteiligten Prozessen ändert sich hier praktisch nichts, lediglich der Ort der Sammlung würde verlagert. In Hinblick auf die von Arbeitsgruppen gepflegten eigenen Lösungen ließe sich hier zusammen mit den Bibliotheken eine nachhaltige Lösung realisieren.

Derzeit existieren einige Ansätze um die heute verwendeten lokalen Bearbeitungswerkzeuge ebenfalls web-basiert zu ersetzen (Perkel 2014). Dies hat v. a. bei verteilten Kollaborationen einige Vorteile. Sollten solche Lösungen globale Autoren-IDs (z. B. ORCID (ORCID 2016)) bereits in den Login-Routinen ermöglichen, was z. B. via OAuth (OAuth Community 2016) mit ORCID (ORCID 2016) als Identity Provider technisch leicht möglich ist, ließen sich aus den Loginkennungen der beteiligten Personen bereits hinreichend gute Metadaten für die Autoren beim späteren Einreichungsprozess gewinnen. Da ferner für diese Systeme ebenfalls Metadaten erforderlich sind (z. B. Titel oder Kurzfassung) können auch diese für den folgenden Einreichungsprozess nachgenutzt werden und eine erneute Erfassung entfielen. Dies sichert neben der Arbeitersparnis eine automatische Konsistenz der Daten.

Da einige dieser Schreibumgebungen bereits heute Zitationen über DOI (via CrossRef) einbetten (z. B. Authorea (Authorea 2016)) ist ein analoger Service der statt CrossRef auf den lokalen privaten Sammlungen eines Repositoriums basiert, und so weitere Dokumentarten (Bücher, Qualifikationsschriften, Vorträge, Daten etc.) erschließt, kein technisches Problem. Repositorien müssen hierzu lediglich den strukturierten Export ihrer Metadaten über eine Web-Schnittstelle erlauben, die sich extern authentifiziert ansprechen lässt. Verstärkt zu adressieren sind in bei Web-basierten Lösungen, vor allem bei

¹⁵ So erlauben JOIN²-Repositorien (JOIN² 2016) bereits den direkten, automatischen Download strukturierter Formate aus manuell erstellten Teilsammlungen, sog. *Körben*. Technisch ließe sich dies über einen einfachen http-get auch nahtlos einbinden.

Cloud-Services, datenschutzrechtliche Fragestellungen. Auch kann *Vertrauen in den Dienst* problematisch sein. Es sollen schließlich keine unfertige Arbeitsversionen der eigenen Publikationen kursieren, bevor die Autoren diese freigeben haben, und die Autoren sollen alle Rechte an ihrer Publikation behalten. Dies kann für eine On-Site Implementation sprechen, und sollte sich in entsprechenden Policies widerspiegeln.

Generell scheint es an dieser Stelle erstrebenswert offene Schnittstellen und Formate zu spezifizieren, um eine breite Akzeptanz zu ermöglichen und ein Vendor Lock-In zu vermeiden. Um eine einfache Adaption zu ermöglichen scheint es geboten, sich stark an derzeit bereits verfügbaren Standards zu orientieren. Kann z. B. eine Schreibumgebung aus einer DOI automatisch eine entsprechende Referenz erstellen, so scheint es sinnvoll einen Export anzubieten, der sich weitgehend an den CrossRef Metadatenschemata (Publishers International Linking Association, Inc. 2016a) orientiert. In der Schreibumgebung wäre dann lediglich eine URL zu ersetzen.

Publizieren: mehr als "Einreichen"

Ist das Dokument geschrieben und liegt die von den Autoren zur Einreichung vorgesehene Fassung vor, steht die Publikation an. Betrachtet man die in der Einleitung beschriebenen *weiteren Verwendungen* genauer, wird schnell ersichtlich, dass es sich hier um den zentralen Punkt des Prozesses handelt, der außerdem das höchste Optimierungspotential bietet:

- Es muss mehr als ein Zielsystem angesprochen werden, und dies ggf. zu verschiedenen Zeitpunkten (z. B. Preprint Server, Repository, Bibliographie/FIS, Verlag)
- Der Eingabeaufwand soll minimiert werden
- Zahlreiche Details sind von den Autoren abhängig und oft nur diesen bekannt (häufig auch nur zum Zeitpunkt des Entstehens der Publikation)
- Die bibliographischen Daten werden vielfach und normiert benötigt, häufig aber erst von Verlagen vervollständigt
- Z. T. werden einrichtungsspezifische Daten benötigt
- Es gelten Randbedingungen aus den Förderrichtlinien
- Datenschutz

Es scheint daher sinnvoll das *Submission Interface* (SI) genauer zu untersuchen. Hierbei soll vom gewünschten Ergebnis ausgegangen werden, und nicht von derzeit realisierten Verfahren, da sich diese für die beschriebenen Anforderungen nur sehr bedingt eignen und ein Festhalten an diesen viele wünschenswerte Verbesserungen verhindert.

Das Submission Interface (SI)

Anforderungen und Potentiale

Diskussionen mit Wissenschaftlern zeigen den Wunsch nach einer Standardisierung des SI. Selbst wenn eine reine Verlagspublikation angestrebt ist, taucht der Wunsch nach einer verlagsübergreifend einheitlichen Maske auf, da niemand z. B. nur in einer einzigen Zeitschrift publiziert und der Einreichungsprozess u. A. durch die verschiedenen Formulare als sehr zeitaufwendig wahrgenommen wird.

Eine Harmonisierung scheint also selbst dann erstrebenswert, wenn keine weiteren Verwendungen angedacht wären. Da eine neue Struktur für alle Publikationen und alle Zielsysteme nutzbar sein soll, muss sie mehr Daten aufnehmen können, als eine heutiges verlagsseitiges Submission Form.

Aus Sicht der Wissenschaftler ist es wünschenswert, dass die notwendigen Eingaben auf ein Minimum reduziert und ggf. notwendige Normierungen transparent erfolgen. Es sollen also nur die für das jeweilige Medium relevanten Daten abgefragt werden. Die Feldauswahl und ggf. die möglichen Inhalte lassen sich über entsprechende Definitionen steuern. Darüber hinaus ist eine maximale Nachnutzung von bereits im Schreibprozess anfallenden Metadaten sinnvoll, was eine Anbindung an die Schreibumgebung wünschenswert macht. So können z. B. Felder wie Titel, Kurzfassung, Volltext automatisch befüllt werden. Erlaubt die *Schreibumgebung* bereits eine Autorenidentifikation so können auch diese Daten übergeben werden. Ein wesentlicher Aspekt an dieser Stelle ist, dass Daten, die in der *Schreibumgebung* nicht strukturiert vorliegen, weiterhin manuell eingegeben werden können und müssen. Dies ist dann nicht besser, *aber auch nicht schlechter* als im derzeitigen Verfahren.

Werden verlagsseitig weitere Daten erforderlich (z. B. Publikationsorgan, Lizenz, Einräumung von Nutzungsrechten) können die dafür möglichen Werte über eine Web-Schnittstelle der Maske verfügbar gemacht und dem Wissenschaftler zur Auswahl angeboten werden.

Gleiches gilt für einrichtungslokale Daten wie z. B. institutionelle Zugehörigkeiten (z. B. Teilinstitute oder Gruppen) oder Projektzuordnungen (lokale Förderschemata, Zuordnung zu Großgeräten etc.). Auch hier können normierte Webservices gültige Werte der Maske zur Verfügung stellen. Je nach Ausgestaltung der *Schreibumgebung* sind diese Daten aber bereits zur nötigen Rechteverwaltung dort vorhanden. Es scheint auch die Anbindung eines *lokalen* Helpdesks sinnvoll, um z. B. auftauchende Fragen zu Finanzierungsmöglichkeiten (die heute häufig erst *nach* der Publikation auftauchen) bereits im Vorfeld klären zu können.

Auch Fördervorgaben könnten über ein solches Interface transparenter transportiert werden. Z. B. hätte auf diese Weise das SI Autoren auf eine vorhandene *Special Clause 39*¹⁶ der FP7-Grants hinweisen können, eine Frage

¹⁶ "[...]In addition to Article II.30.4, beneficiaries shall deposit an electronic copy of the published version or the final manuscript accepted for publication of a scientific publication relating to foreground published before or after the final report in an institutional or subject-based repository at the moment of publication.[...]"(European Commission 2012)

die aktuell häufig erst im Kontext der Abschlussberichte auftaucht¹⁷. Die Maske könnte ebenfalls automatisch auf entsprechende Publikationsfonds verweisen (im Beispiel ein Refunding durch die EU, oder auch ggf. verfügbare Fonds innerhalb eines DFG-Projekts oder der Einrichtung), wenn der Verlag APCs oder andere Gebühren geltend macht¹⁸. Ein entsprechendes Budget könnte bereits an dieser Stelle blockiert werden¹⁹. Ist das zu belastende Konto bereits im Einreichungsprozess definiert, sind die Kosten für die Einrichtung und die Förderer kalkulierbar und es eröffnet sich die Möglichkeit für entsprechende Etatverteilungsmodelle (analog zu z. B. (Trott 2004)), so dass auch *Orchideenfächer* keine Sorge haben müssen ihre Publikationen zum Jahresende nicht mehr finanzieren zu können. Gleichzeitig ermöglicht dies den Wissenschaftlern den einfachen Umgang mit diesen neuen Instrumenten.

Auch Verlage würden von einer solchen Harmonisierung des *SI* stark profitieren. Systemimmanent wäre die Kostendeckung bereits *vor* Einreichung der Publikation final geklärt. Da die nötigen Fakturierungsdaten aus den jeweiligen relevanten Verwaltungssystemen zur Verfügung gestellt werden und die Deckung garantiert ist, kann von einer stets korrekten Rechnungslegung und einer mithin zügigen Bearbeitung ausgegangen werden. Die Teilung von Publikationsgebühren zwischen beteiligten Einrichtungen würde ebenfalls bereits vor Einreichung festgelegt und deren Kostendeckung geprüft.

Da ggf. nötige Kommunikation hier mit zentralen Einrichtungen stattfindet verbliebe man bei dem bisher im Bereich *Subskription* vorhandenen Modell weniger Experten als Partner bei der Abwicklung. Eine Umstellung von Einzelrechnungen auf Sammelabrechnungen wie auch in (Bruch u. a. 2015) angeregt wäre damit deutlich einfacher realisierbar. Der Verwaltungsaufwand sowohl auf Verlags- als auch auf Einrichtungsseite würde sich signifikant reduzieren.

¹⁷ Erfreulicherweise führt dies zumindest am DESY derzeit einer retrospektiv ansteigenden Open Access-Quote aus diesen Programmen, da über diese Vorgabe zumindest vom Zweitveröffentlichungsrecht bzw. dem Grünen Weg gebraucht gemacht wird. Allerdings hatte sich der Förderer sicherlich eine zeitnähere Open Access-Publikation erhofft.

¹⁸ Die in (Bruch u. a. 2015) geforderte Einführung einer zentralen Stelle zur Verwaltung von Publikationsgebühren setzt ein derartiges Verfahren praktisch voraus, da sich diese zentralen Fonds nach derzeitigem Stand aus verschiedenen, z. T. zweckgebundenen, Quellen speisen werden.

¹⁹ Es *kann* und *soll* nicht von Wissenschaftlern erwartet werden, dass sie sich gleichzeitig zu Verwaltungsfachleuten entwickeln, die diese gesamte komplexe Gemengelage im Detail überblicken, insbesondere da die hierfür nötigen Verwaltungsstrukturen bereits alle existieren.

Anders als in bisherigen Prozedere, bei denen *nach* der Publikation aus i. d. R. unnormierten Freitexten mit Affiliationsangaben oder gar der Danksagung die Einrichtungen und Förderer mit viel Aufwand *ihre* Publikationen ermitteln (wobei systemimmanent nur diejenigen ermittelt werden können, die in anderen Datenbanken nachgewiesen sind) stünde im vorgeschlagenen Verfahren die Affiliationszuordnung bereits bei Einreichung fest. Die Forderung an Verlage diese Daten strukturiert zu liefern, ist damit obsolet. Das SI würde nicht nur das Verlagssystem mit Daten versorgen, sondern die Daten um die vom Verlag vergebenen Werte (z. B. DOI, Zeitschrift, Volume, Jahr, Seiten etc.) ergänzt direkt der Bibliographie und einem ggf. angeschlossenen FIS verfügbar machen. Gleiches gilt für das Einpflegen des Volltextes in die institutionellen Repositorien²⁰. Es ist hervorzuheben, dass in diesem Verfahren verlagsseitig *keine* zusätzlichen Arbeiten anfallen. Reicht momentan der eigene Webserver die Formulardaten an die Verlagsdatenbank weiter, würde dies künftig über ein normiertes Format ggf. von einem anderen Host aus erfolgen.

Vom Monolith zum Web-Service

Da es beim Einsatz von Web-Technologien nicht relevant ist, wo das SI realisiert wird, ist hier die Definition von offenen, normierten Datenformaten entscheidend. Die Übertragung kann dann über heutige Web-Technologien erfolgen. Da zahlreiche Werte einrichtungsinterne vorliegen, scheint eine Ansiedlung des SI an den Forschungseinrichtungen selbst sinnvoll. Dies kommt u. U. auch den Autoren entgegen, die der eigenen Einrichtung häufig mehr vertrauen als einem fremden System.

Auch ein entsprechendes Branding bzw. die Anpassung der Feldbezeichnungen an lokale Gegebenheiten (z. B. gliedern sich manche Institutionen in Institute, andere in Gruppen oder Departments) scheint auf diese Weise einfacher und käme den Benutzern entgegen.

Da aber für eine Publikation im Allgemeinen einrichtungsinterne Daten *verschiedener* Institutionen relevant sind²¹ und daher die Befüllung der

²⁰ Da die nötigen Schnittstellen hierbei aus Sicht des SI immer 1-n ausgelegt sind ist auch der Fall abgedeckt, dass mehrere Institutionen gemeinsam publizieren oder das FIS-Modul, das sich um Publikationen kümmert, noch nicht das Repository darstellt.

²¹ Man denke an die Kollaborationen von Autoren verschiedener Einrichtungen.

jeweiligen Vorschlagslisten über Webschnittstellen erfolgen muss, ist eine Ansiedlung an der eigenen Einrichtung nicht strikt notwendig. Bei genauer Betrachtung kapselt das S/ aus systemischen Gründen den Prozess an sich in einem strikt privaten Bereich. So kann man sich durchaus vorstellen, dass zur Einreichung das S/ irgendeiner Einrichtung verwendet wird, selbst einer, die an der konkreten Publikation nicht beteiligt ist²². Selbst eine Auslegung als verteiltes System, welches den zu verwendenden Host zufällig auswählt ist denkbar. Neben entsprechenden Richtlinien ist dies ein wirksames Mittel, datenschutzrechtlichen Vorbehalten entgegenzutreten. Per se haben zwar Einrichtungen oder Förderer am Procedere des Einreichungsvorgangs kein Interesse, es könnte aber der Eindruck entstehen, dass sich z. B. Ablehnungen von Publikationen negativ auf Bewertungen auswirken, wenn diese der Einrichtung bekannt sind²³.

In allen diesen Fällen sammelt das S/ zunächst an zentraler Stelle Metadaten, sowie die für die Submission vorgesehenen Volltexte und legt diese in einem temporären Speicher ab²⁴.

An dieser Stelle sollten bereits die für spätere Verwendungen nötigen zusätzlichen Metadaten wie Projektzuordnungen²⁵ oder Affiliationen²⁶ erfasst werden, da diese i. d. R. nur den Autoren bekannt sind.

Es ist ferner zu berücksichtigen, dass Publikationen im S/ optional noch Genehmigungsprozessen innerhalb einer Institution unterliegen können, bevor diese zur Publikation freigegeben werden. Z. B. müssen in den meisten

²² Möchte man trotzdem an einem einrichtungsspezifischen Branding oder Wording festhalten müssen diese Parameter über entsprechende Webschnittstellen transportiert werden.

²³ Allerdings wäre das natürlich nicht sachgerecht, da ein üblicher Ablehnungsgrund ist, dass der Herausgeber einen Artikel als zu weit vom Kernbereich der Zeitschrift entfernt ansieht (*out of scope*). Hier wird also aus eher formalen denn aus wissenschaftlichen Gründen abgelehnt. Überhaupt ist die Bewertung der Qualität von Aufsätzen auf Basis der Zeitschriften in denen diese erscheinen fragwürdig.

²⁴ Einige Repository-Lösungen bieten hierfür bereits fertige Lösungen z. B. der HoldingPen in invenio (Invenio Development Team 2016).

²⁵ Eine mögliche Quelle wäre z. B. FundRef (Publishers International Linking Association, Inc. 2016b). Allerdings müsste hier die Granularität noch deutlich ausgebaut und die Lieferung der relevanten Grants durch die Förderer sichergestellt werden. Da hier allerdings immer Web-Schnittstellen relevant sind kann hier auch eine Anbindung an interne Projektdatenbanken der an einer Publikation beteiligten Institutionen erfolgen. (Große 2014)

²⁶ Derzeit fehlen noch hinreichend granulare und frei zugängliche Datenbanken ausreichender Qualität, um diesen Teil abzubilden. Da sich Affiliationen, insbesondere auf Teilinstitutsebene, häufig ändern scheint es, bis automatische normierte Datenlieferungen an einen zentralen Hub erfolgen können sinnvoller, diese Werte aus den internen Daten der beteiligten Einrichtungen zu beziehen. Da auch hier immer Web-Schnittstellen relevant sind ist ein zentraler Host aber wie bei Projekten *keine* Voraussetzung. (Große 2014)

Einrichtungen der Helmholtz-Gemeinschaft die jeweiligen Instituts-/Gruppenleiter als Vorgesetzte die Publikation an sich genehmigen²⁷.

Um ein solches System zu realisieren sollte ein Baukastenprinzip angestrebt werden. Liegen Web-Technologien zu Grunde ist dies vergleichsweise einfach realisierbar. So ist es möglich nach und nach neue Funktionen zu integrieren und sich flexibel an neue Aufgaben anzupassen.

So wäre aus heutiger Sicht zunächst eine Abstraktion der vorhandenen verlagsspezifischen Schnittstellen ein Fortschritt. Dies würde zunächst lediglich die Definition des Datenaustauschformats sowie die Auslegung von Login-Prozeduren und der verwendeten Protokolle erfordern. In einem ersten Schritt würde sich also, bis auf die technische Basis nicht viel ändern.

In weiteren Ausbaustufen können zusätzliche Module eingebaut werden, um z. B. eine Autorenidentifikation via ORCID (ORCID 2016), die Anbindung von FundRef (Publishers International Linking Association, Inc. 2016b) und weiterer Normdatensysteme zu realisieren.

Da aus jetziger Sicht weder ORCID (ORCID 2016) noch FundRef (Publishers International Linking Association, Inc. 2016b) geeignet sind die Anforderungen zu erfüllen, schafft dies auch die Möglichkeit nach und nach Verbesserungen an den angeschlossenen Systemen vorzunehmen. Die Herausforderungen liegen hier eher in der Datenqualität als im technischen Aufsetzen entsprechender webprotokoll-basierter Schnittstellen. Es sei hervorgehoben, dass in den derzeit diskutierten Modellen solche Anbindungen an die Verlage als Forderungen herangetragen werden, so dass jeder Verlag solche Schnittstellen in Eigenregie erstellen muss. Es scheint daher auch im verlagsinteresse, wenn hier eine abstrakte, allgemeine Lösung geschaffen wird, die als OpenSource nachgenutzt werden kann.

Sollten sich, was zu erwarten ist, mittelfristig web-basierte Systeme zum Erstellen von Publikationen etablieren, kann als nächster Schritt eine Schnittstelle zu diesen Systemen geschaffen werden, die in zunächst nur die Übernahme der Volltexte erlaubt und sukzessive um die Übergabe weiterer Metadaten je nach

²⁷ Gründe hierfür sind z. B. Wahrung von Patentmöglichkeiten, Einhaltung guter wissenschaftlicher Praxis, Haftung etc. Man kann diese Prozesse natürlich auch außerhalb des SI abbilden. Sie weisen aber strukturell hohe Ähnlichkeiten mit Reviews und der Lieferung von Referee Reports auf.

Fähigkeit des angeschlossenen Systems erweitert wird (Titel, Kurzfassung, Autoren, Affiliationen etc.). Auf Seiten des *SI* sollte vorgesehen sein, dass fehlende Daten manuell ergänzt werden bis sie dem zu definierenden Austauschformat entsprechen. Die Anzahl der Pflichtfelder ist hier, um eine einfache Anbindung zu ermöglichen, weitestgehend zu reduzieren.

Bereits heute steht mit SWORD (swordapp.org 2016) ein Protokoll zur Verfügung das nach dem *Push*-Prinzip wesentliche Teile für die Übertragung eines Volltexts aus einer Schreibumgebung an ein *SI* zur Verfügung stellt. Auch zahlreiche Versionskontrollsysteme wie *git* stellen Möglichkeiten zur Übertragung zur Verfügung. Definiert man die Metadaten in einer entsprechenden Datei (Filialdateien oder sidecar files (Wikipedia-Autoren 2015)) ist auch eine solche Anbindung mit heutigen Mitteln einfach möglich, sofern das Datenschema und -format geklärt sind.

Eine alternative Implementation würde statt des *Push* aus der Schreibumgebung ein *Pull* von diesem System realisieren, also einen *Datenimport* in das *SI*. Dies ist analog dem *harvesten* eines Einzeldokuments, so dass OAI-PMH hier bereits heute die nötigen Funktionen realisiert. Ein solches Verfahren setzt allerdings web-basierte Schreibumgebungen voraus, während man bei Push-Schnittstellen auch eigenständige Clients realisieren kann, die auf lokalen Dateien arbeiten. Solche eigenständigen Clients könnten ggf. auch die Umsetzung lokaler Autorenidentifikationssysteme (Logins, Email-Adressen) in globale Systeme wie ORCID (ORCID 2016)) übernehmen. Die Ausgestaltung und der Umfang der Daten läge hier auf Seiten des Clients.

Soll die Publikation schlussendlich eingereicht werden leitet das *SI* die dort gesammelten Metadaten weiter. Im Falle der Publikation auf einem Repositorium entspricht dies einem einfachen Deposit. Zusätzlich soll der Autor die Möglichkeit haben weitere Repositorien (z. B. arXiv) zur Einreichung auszuwählen. Diese bekämen ebenfalls den normierten Metadatensatz und könnten ihn übernehmen. Ist eine Publikation in einem Verlag angestrebt, soll das *SI* hier das jeweilige Organ zur Auswahl anbieten, und über eine Web-Schnittstelle die jeweils geltenden Bedingungen des Verlags für den Autor darstellen. Dies versetzt den Wissenschaftler in die Lage an dieser Stelle eine informierte Entscheidung zu treffen und im Sinne des Marktgedankens Angebote abzuwägen und auszuwählen.

Ist die Publikation kostenpflichtig würde dies in einer ersten Stufe lediglich mit allen anfallenden Kosten aufgelistet und angezeigt. Mit weitergehender Automatisierung könnte hier die Deckung über lokale Fonds geprüft werden. Im einfachsten Falle würde lediglich auf den entsprechenden Helpdesk verwiesen, und die Kostenübernahme manuell durch den Autor bestätigt. Sodann wird das Dokument dem Verlag übermittelt, der dieses ggf. in den jeweils üblichen Prozess übernimmt. Entsprechende Antworten können den Autoren ebenfalls über das *SI*/verfügbar gemacht werden. Hierzu müssen diese nicht in das *SI* übertragen werden, sondern können online vom jeweiligen Verlag abgerufen werden. Dies stellt sicher, dass das *SI* keine Bewertungen des jeweiligen Aufsatzes kennt, was aus Datenschutzgründen sinnvoll sein kann. Soll nach einer endgültigen Ablehnung die Publikation ggf. nach einer Überarbeitung bei einem anderen Organ eingereicht werden, kann dies ohne Neueingaben lediglich durch eine neue Auswahl geschehen.

Damit Überarbeitungen einfach möglich sind ist eine möglichst gute Anbindung an die Schreibumgebung hilfreich. Im einfachsten Fall erlaubt diese nur Austausch des Volltextes durch eine neue Version, aber auch eine komplette Versionierung inkl. der Aktualisierung der Metadaten ist denkbar. Realisiert mit eindeutigen Identifiern für die Publikation und über offene Webprotokolle separiert dies auch die Schreibumgebung vom *SI*.

Bei Annahme der Publikation übernimmt das *SI* die automatische Speicherung der endgültigen Fassung die ggf. notwendige Aktualisierung der bereits im initialen Prozess angestoßenen Deposits auf Fachrepositorien durch Übermittlung des finalen, normierten Metadatensatzes sowie auf dem jeweiligen Repositorium. Fallen für die Publikation Kosten an können die jeweiligen Modalitäten direkt mit korrekten Daten an den jeweiligen Verlag weitergereicht werden, so dass je nach Verlagsprozess eine entsprechende Rechnungslegung mit korrekten Daten erfolgt.

Durch Übernahme in das eigene Repositorium steht die eigene Publikation ferner für neue Arbeiten zur Zitation zur Verfügung. Dass das *SI* auch die Einreichungsformen des Dokuments bis hin zum *post print* kennt, lässt sich auch der *Grüne Weg* für den Autor einfach realisieren. Sind mit den Verlagen z. B. im Kontext der *Allianz-Lizenzen* zusätzliche Open Access-

Vereinbarungen getroffen können diese an dieser Stelle, verbunden mit der damit einher gehenden Steigerung der Visibilität, bereits realisiert werden.

Da das *SI* bereits bei Einreichung, also *vor* dem Entstehen der finalen Publikation, die entsprechenden Daten erfasst ist eine Vollständigkeit prinzipiell möglich. Ferner ist an dieser Stelle bereits die Affiliationszuordnung per Prozess geklärt, eine Herausforderung die in fast allen anderen Kontexten derzeit als unmöglich gilt²⁸.

Ist das Repositorium der Einrichtung nicht gleichzeitig die jeweilige Publikationsdatenbank bzw. der für Publikationsdaten zuständige Teil eines FIS(Große 2014) können die normierten Daten auch hier direkt deponiert werden. Dies stellt größt mögliche Konsistenz und Vollständigkeit der Meldungen im Sinne der Wissenschaft sicher²⁹. Über die Normierung der Depositschnittstellen und Metadatenformate können alle beteiligten Institutionen von einer Stelle aus versorgt werden. Als gewünschter Nebeneffekt sorgt die automatische Anbindung an die jeweiligen institutionellen Repositorien für eine entsprechende Indexierung durch Suchdienste wie z. B. BASE (Universität Bielefeld 2016) und erhöht so die Sichtbarkeit.

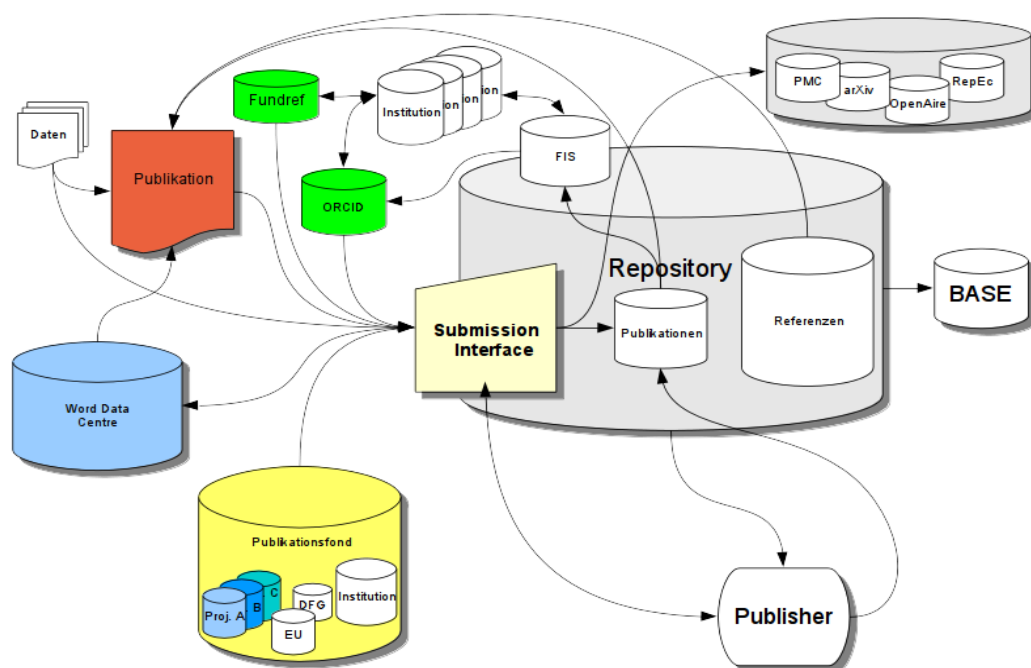
Heute ist nach der Publikation vor dem Bericht, d. h. es müssen, je nach Bericht z. T. extrem aufwendig, alle Publikationen einer Einrichtung, eines Projekts, einer Person zusammengestellt, und dafür heute i. d. R. zunächst erfasst werden (Große 2014). Diese aufwendige Sammlung entfällt im vorgeschlagenen Modell, da das *SI* automatisch die entsprechenden Systeme beliefern kann. Projektzuordnungen müssen nicht aus textuellen Danksagungen intellektuell ermittelt werden, sondern liegen bereits ab Einreichung strukturiert vor. Dies ermöglicht den beteiligten Wissenschaftlern ein korrektes Reporting auch wenn ein Projekt mehrere Jahre läuft. Ebenso ermöglicht sie den Förderern mit Hilfe der normierten Projektzuordnungen Publikationen zuzuordnen selbst wenn diese erst nach dem Projektende final vorliegen. Verknüpft man daher Meldung mit der Einreichung spart dies

²⁸ Vgl. den wesentlich aufwendigeren Prozess(Koch u. a. 2016), der ohne eine solche Schnittstelle notwendig ist insbesondere die diskutierte Problematik der dort relevanten *retrospektiven* Affiliationszuordnung.

²⁹ Ist die Schreibumgebung das *übliche* Werkzeug zum Erstellen von Dokumenten können hier auch Lehrmaterialien wie z. B. Vorlesungsskripte direkt verfügbar gemacht werden. Es findet lediglich keine Übergabe an einen Verlag statt, sondern die Publikation über das lokale Repository.

einerseits die erneute Erfassung und sichert andererseits eine zeitnahe Verfügbarkeit konsistenter Daten an den gewünschten Stellen. Dies gilt auch für die eingangs erwähnte Außendarstellung.

Eine schematische Darstellung der jeweiligen Datenflüsse könnte wie folgt aussehen:



Möglicher Datenfluss und Verbindungen von Systemen, die im Kontext einer wissenschaftlichen Publikation relevant sind.

Schlussfolgerungen

In der Vergangenheit dienten bibliographische Daten im Wesentlichen zum Austausch von Informationen in einer kleinen Fachöffentlichkeit. Sie dienten dazu, dass Wissenschaftler relevante Publikationen in ihrem eigenen Forschungsfeld leicht finden konnten und die dortigen Ergebnisse in ihren eigenen Arbeiten verwenden zitieren konnten. Hierfür sind vergleichsweise wenige Daten erforderlich, zumal je nach Feld die Anzahl relevanter

Publikationsorgane übersichtlich war und die Zuordnung immer intellektuell erfolgte.

Neben diese klassische Verwendung bibliographischer Daten treten zunehmend weitere Anwendungen. So hat heute auch die allgemeine Öffentlichkeit Interesse an den Ergebnissen der Forschung. Durch den gestiegenen Wettbewerb in der Wissenschaft wird ferner eine Darstellung der eigenen Leistung immer wichtiger.

Dies gilt sowohl für den einzelnen Wissenschaftler als auch für entsprechende wissenschaftliche Einrichtungen, die sich alle im Wettbewerb um endliche Ressourcen befinden. Da hier außerdem überwiegend automatische Verfahren zum Einsatz kommen sollen, steigen die Anforderungen an die vorhandenen Metadaten und deren Qualität deutlich.

Vor diesem Hintergrund scheint es geboten den Prozess der Publikation zu überdenken. Zusammen mit den heute im Web verfügbaren Technologien können hier neue Überlegungen bisherige Verfahren ersetzen und einen entsprechenden Mehrwert für alle beteiligten Schaffen. Insbesondere die heute verwendeten *SI*, die lediglich die Systeme der Verlagshäuser beliefern, erscheinen nicht mehr zeitgemäß.

Es wird daher vorgeschlagen, das eigentliche *Submission Interface* abstrakter zu gestalten und auf offenen Protokollen und Datenformaten basieren zu lassen. Dies würde erlauben, Publikationen bereits beim Einreichen mit den notwendigen Metadaten zu versehen, die es erlauben die verschiedenen Anforderungen zu erfüllen, die heute an bibliographische Daten gestellt werden (Institut für Forschungsinformation und Qualitätssicherung, Fraunhofer-Institut für Angewandte Informationstechnik und Wissenschaftsrat, Geschäftsstelle 2015). Damit können die aktuell noch notwendigen Mehrfacherfassungen entfallen. Die gemachten Vorschläge hierzu basieren ausschließlich auf heute bereits vorhandenen Technologien. Der vorgeschlagene Prozess erlaubt eine sukzessive Überführung des bestehenden Systems, wobei mit möglichst geringen Mitteln Vorteile für alle Beteiligten (Wissenschaftler, Verlage, Einrichtungen, Förderorganisationen) realisierbar werden.

Bibliographie

Authorea. 2016. Authorea. <https://www.authorea.com>.

Björk, Bo-Christer und David Solomon. 2014. Developing an Effective Market for Open Access Article Processing Charges. <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Guides/WTP054773.htm>.

Bruch, Christoph, Gernot Deinzer, Kai Geschuhn, Petra Hätscher, Kristine Hillenkötter, Ulrike Kreß, Heinz Pampel, u. a. 2015. Positionen zur Schaffung eines wissenschaftsadäquaten Open-Access Publikationsmarktes : Positionspapier der ad-hoc-AG Open-Access-Gold im Rahmen der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. Potsdam: Ad-hoc-Arbeitsgruppe Open-Access-Gold der Schwerpunktinitiative „Digitale Information“ der Allianz der deutschen Wissenschaftsorganisationen. doi:10.2312/allianzoa.008,.

Cornell University Library. 2016. ArXiv.org. <http://arxiv.org>.

Dropbox. 2016. Dropbox. <https://dropbox.com>.

Dublin Core Metadata Initiative. 2016. Dublin core metadata element set, version 1.1. <http://dublincore.org/documents/dces/>.

European Commission. 2012. *Special clause 39*. http://ec.europa.eu/research/press/2008/pdf/annex_1_new_clau.

---. 2016a. Horizon 2020 - The EU Framework Programme for Research and Innovation. <https://ec.europa.eu/programmes/horizon2020>.

---. 2016b. OpenAIRE. <http://openaire.eu>.

Ewert, Gisela und Walther Umstätter. 1997. *Lehrbuch der Bibliotheksverwaltung*. Hg. von Wilhelm Krabbe und Wilhelm Martin Luther. [4. Aufl.]. Stuttgart: Hiersemann.

Google, Inc. 2016a. Google Drive. <https://www.google.com/drive/about>.

---. 2016b. Google Docs. <https://www.google.com/docs/about>.

Große, Katrin. 2014. Forschungsevaluation aus dem Repositorium. In: OpenAccess Tage 2014, Köln (Germany), 8 Sep 2014 - 9 Sep 2014, 8. September. <http://repository.gsi.de/record/67551>.

Institut für Forschungsinformation und Qualitätssicherung, Fraunhofer-Institut für Angewandte Informationstechnik und Wissenschaftsrat, Geschäftsstelle. 2015. Spezifikation des Kerndatensatz Forschung - Version 1.0. <https://bib-pubdb1.desy.de/record/296180>.

Invenio Development Team. 2016. Invenio. <http://inveniosoftware.org>.

JOIN². 2016. JOIN² - just anOther iNvenio iNstance. <http://join2.de>.

Koch, Thorsten, Rolf Griebel, Konstanze Söllner, Jürgen Christof und Roland Bertelmann. 2016. DeepGreen - Entwicklung eines rechtssicheren Workflows zur effizienten Umsetzung der Open-Access-Komponente in den Allianz-Lizenzen für die Wissenschaft. ZIB. <http://nbn-resolving.de/urn:nbn:de:0297-zib-56799>.

OAuth Community. 2016. OAuth. <http://oauth.net>.

OrbiTeam Software GmbH & Co. KG. 2016. BSCW. <https://public.bscw.de/pub/>.

ORCID. 2016. ORCID. <http://orcid.org>.

ownCloud. 2016. OwnCloud. <https://owncloud.org/>.

Perkel, Jeffrey M. 2014. Scientific writing: The online cooperative. *Nature* 514, Nr. 7520 (Oct): 127–128. doi:10.1038/514127a,.

Publishers International Linking Association, Inc. 2016a. Crossref deposit schema. http://help.crossref.org/deposit_schema.

---. 2016b. FundRef. <http://www.crossref.org/fundingdata>.

swordapp.org. 2016. Sword. <http://swordapp.org/about>.

The Etherpad Foundation. 2016. Etherpad. <http://etherpad.org/>.

Trott, Sabine. 2004. *Entwicklung von Erwerbsprofilen für die Fächer Mathematik und Physik an der UB Ilmenau*. Bd. 133. Berliner handreichungen zur bibliothekswissenschaft. Institut für Bibliothekswissenschaft der Humboldt-Universität zu Berlin. <http://www.ib.hu-berlin.de/~kumlau/handreichungen/h133/>.

Universität Bielefeld. 2016. BASE - bielefeld academic search engine. <https://www.base-search.net>.

Wagner, Alexander. 2013. JuSER – Bibliographie, OpenAccess-Server, Literaturverwaltung... In: *Zentrale hochschulbibliothek luzern*. Luzern. <http://hdl.handle.net/2128/5470>.

---. 2016. JOIN² - What it is and what it does. In: Privatuniversität Witten/Herdecke. <https://bib-pubdb1.desy.de/record/296178>.

Wikipedia-Autoren. 2015. Filialdateien. *Wikipedia*. <https://de.wikipedia.org/w/index.php?title=Filialdatei&oldid=144778980>.

Wissenschaftsrat. 2016. Empfehlungen zur Spezifikation des Kerndatensatz Forschung. Wissenschaftsrat. http://www.wissenschaftsrat.de/arbeitsbereiche-arbeitsprogramm/kerndatensatz_forschung.html.